

ESTIMATING REDSHIFT DISTRIBUTIONS WITH SPATIAL CORRELATIONS: METHOD AND APPLICATION TO DATA

BRICE MÉNARD^{1,2,3}, RYAN SCRANTON⁴, SAMUEL SCHMIDT⁴,
 CHRIS MORRISON⁴, DONGHUI JEONG¹, TAMAS BUDAVARI¹, MUBDI RAHMAN¹

Draft version March 20, 2013

ABSTRACT

We present a method to infer the redshift distribution of an arbitrary dataset based on spatial cross-correlation with a reference population and we apply it to various datasets across the electromagnetic spectrum. Our approach advocates the use of clustering measurements on all available scales, in contrast to previous works focusing only on linear scales. We apply this technique to infer the redshift distributions of luminous red galaxies and emission line galaxies from the SDSS, infrared sources from WISE and radio sources from FIRST. We show that consistent redshift distributions are found using both quasars and absorber systems as reference populations. This technique promises to be widely applicable to existing and upcoming sky surveys. It provides us with the invaluable ability to deproject the inherently 2d observations of the extragalactic sky.

Subject headings: redshift – clustering

1. INTRODUCTION

Observations of the sky are inherently a two-dimensional measurement of photon flux density as a function of angular position. For astrophysical studies one usually needs to infer three-dimensional positions, for example to convert a brightness into a luminosity. This has been a long-standing limitation in astronomy.

On extragalactic scales, distances are usually inferred from redshift measurements using the knowledge of the expansion history of the Universe. A redshift can be directly measured from observations when one can detect and identify a high-contrast spectroscopic feature. Consequently, robust redshift measurements require spectroscopic observations of sources with emission or absorption lines at a sufficient spectral resolution. Such observations are usually expensive and restricted to bright objects; for example, the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) has imaged ~ 100 million galaxies, but only of order 1% have been followed-up spectroscopically, most of which are bright and nearby.

For the vast majority of galaxies, distance estimates rely on so-called “photometric” redshifts. They use observed broadband colors to probe the overall spectral energy distribution (SED) of a source. Thus, they rely of qualitatively different information. Photometric redshift estimation suffers from a number of limitations: intrinsic degeneracies between colors and redshifts, arbitrary SED templates, dust reddening, etc. Despite these limitations, however, all upcoming imaging surveys rely on photometric redshifts. With deeper surveys of the sky and access to new wavelength ranges from space, the lack of robust distance estimates is becoming a limitation. Moreover, given the speed that telescopes

are mapping out the sky, the fraction of objects for which we have spectra *decreases* with time. Consequently, alternative techniques should be explored to estimate cosmological redshifts.

In this paper, we show that the distance estimates from the tiny fraction of sources with spectroscopic or accurate photometric redshifts can be propagated statistically to other objects across different surveys using information extracted from spatial clustering. The main idea is that the large-scale cross-correlations between objects with existing distance measurements and an unknown astronomical sample can provide us with an estimate of the unknown sample’s redshift distribution.

While this direction has been explored previously by several authors (Newman 2008; Benjamin et al. 2010; Matthews & Newman 2010; Schulz 2010; Matthews & Newman 2012; McQuinn & White 2013), in this paper we show that this idea can be generalized by including clustering information from all scales, leading to a much higher sensitivity and a wider applicability. We demonstrate the power of this technique by estimating the redshift distributions of existing datasets, without any knowledge of the source properties. We apply our tool to various datasets across the electromagnetic spectrum, from the optical to the radio range (where photometric redshifts cannot even be defined) and estimate the corresponding redshift distributions. A companion paper (Schmidt et al. 2013) presents results from numerical simulations to test the robustness and limits of our redshift inference method when applied to realistic distributions of dark matter halos and galaxies.

2. ESTIMATING REDSHIFTS

2.1. The covariance of the sky

Electromagnetic observations of the sky consist of a measurement of flux density F_λ as a function of angular position. We denote the flux density fluctuation at a location ϕ and wavelength λ as $\delta F_\lambda(\phi)$. The generic

¹ Department of Physics & Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA

² Institute for the Physics and Mathematics of the Universe, Tokyo University, Kashiwa 277-8583, Japan

³ Alfred P. Sloan Fellow

⁴ Department of Physics, University of California, One Shields Avenue, Davis, CA 95616, USA

covariance of the extragalactic sky is given by

$$C_{\text{obs}}(\lambda_1, \lambda_2, \theta) = \langle \delta F_{\lambda_1}(\phi) \delta F_{\lambda_2}(\phi + \theta) \rangle_{\phi}. \quad (1)$$

This quantity is uniquely defined and provides us with the statistical information on the extragalactic sky as a function of position and wavelength. Given that the large-scale structure of the Universe approaches that of a Gaussian random field, this quantity captures a significant fraction of the structure of the extragalactic sky.

If we have access to a population of objects whose spatial distribution $\delta(\vec{r})$ is located within a narrow redshift bin centered on z , i.e. in the limit of $dN/dz \rightarrow \delta_D(z' - z)$, we can use it to probe a projection of the flux density fluctuation δF_{λ} :

$$C_{\text{obs}}(\lambda, r_p, z) = \langle \delta(z) \delta F_{\lambda}(r_p) \rangle. \quad (2)$$

If this correlation is significantly greater than random, it implies that the field δF contains sources at redshift z (we first ignore the small modulation possibly induced by gravitational magnification and discuss it in section 2.4). Here we note that the flux fluctuation δF_{λ} can be a continuous field (e.g. CMB temperature) or a discrete one (e.g. galaxies). This indicates that it is possible to extract statistical information on the redshift distribution of an arbitrary dataset from measurements of this cross-correlation as a function of redshift. To first order, the correlation in Equation 2 can be used directly to test for the absence of objects in δF_{λ} at redshift z .

When using observed spectra to calibrate photometric redshifts, one makes use of the correlation given in Eq. 2, the correlation between a known redshift and the observable δF_{λ} (or similarly a color), but ignores the spatial dependence. An important point of this paper is that the environment (or projected environment) of a source can be treated as an observable which, in a statistical context, is a powerful indicator of its properties, including its redshift. Due to overlapping objects along the line-of-sight, the projected environment is often a noise-dominated quantity. However, if one is interested in estimating the redshift of an ensemble of objects, the mean projected environment can become a signal-dominated quantity and a useful source of information. We now show how to make use of this information to infer a redshift distribution for the ensemble of objects. A few authors have explored this avenue (Newman 2008; Matthews & Newman 2010; Schulz 2010; Matthews & Newman 2012; McQuinn & White 2013, e.g.) and considered approaching the problem from a theoretical and global point of view. Here we take a different approach, taking into account its applicability to real-world datasets, emphasizing the interest of using small scale clustering information.

2.2. Redshift inference from spatial clustering

Let us consider two populations of extragalactic objects: a *reference* population for which we know the angular positions and redshifts of each object and an *unknown* population for which angular positions are known but redshifts are not. Let n_i be the surface density of objects and δn_i be the corresponding density fluctuation $\delta_i = n_i / \langle n_i \rangle - 1$, where $\langle n_i \rangle$ is the mean density. The index i refers to the reference or unknown population.

We introduce the normalized redshift distribution:

$$\phi_i(z) = \frac{d^2 N_i}{dz d\Omega} \bigg/ \int dz d\Omega \frac{d^2 N_i}{dz d\Omega} \quad (3)$$

where $d^2 N_i / dz d\Omega$ is the number of objects per unit redshift and solid angle. The clustering of matter induces correlations between the positions of overdensities. The mean density of unknown sources at a given separation from reference objects is given by

$$\langle n_u(\theta, z) \rangle_r = \langle n_u \rangle [1 + w_{ru}(\theta, z)] \quad (4)$$

where $w_{ru}(\theta, z)$ is the angular cross-correlation function between *all* unknown objects and the reference population at redshift z (Peebles 1993). This is the basic source of information we will use to infer redshifts.

Previous studies making use of spatial correlations to infer redshift distributions focused only on large scales, on which the galaxy bias is sufficiently linear. For example, Newman (2008) proposed a method to recover the redshift distribution of unknown populations, using measurements of the cross-correlation function on scales greater than a few Mpc. Here, instead we propose to measure the amount of clustering by integrating over all available scales. This approach allows us to significantly increase the signal-to-noise ratio (S/N) of the basic observable. From this quantity alone, valuable information can be extracted; It can, for example, be used to infer the existence or absence of sources at a given redshift. As we will show below, it can also be used to probe the local properties of an unknown redshift distribution. While somewhat less accurate than the method proposed by Newman (2008) and recently optimized by McQuinn & White (2013), our proposed method is significantly more sensitive and can be applied to numerous datasets. From a practical point of view we also note that analyses of real-world data (for example in the optical and the infrared regimes) tend to be much less affected by systematics on small scales. Working on scales smaller than a degree (or a few projected Mpc) can be quite valuable. In this paper, we advocate for this approach and demonstrate that cluster-based redshifts provide us with a powerful exploratory tool.

As a measure of clustering we will consider the integrated cross-correlation function

$$\bar{w}_{ur}(z) = \int_{\theta_{\min}}^{\theta_{\max}} d\theta W(\theta) w_{ur}(\theta, z) \quad (5)$$

where $W(\theta)$ is a weight function, whose integral is normalized to unity, aimed at optimizing the overall S/N. As the matter correlation function can often be approximated by a power law over a broad range of scale with γ or order unity, we can simply use $W(\theta) \propto \theta^{-1}$. We note that for $\gamma = 1$ there is an equal amount of clustering information per logarithmic scale. This suggests that a significant amount of information can be extracted from small scale measurements.

The integrated angular cross-correlation between the *reference* sample and the *unknown* sample can be written as

$$\bar{w}_{ur} = \int dz' \phi_u(z') \phi_r(z') \bar{b}_u(z') \bar{b}_r(z') \bar{w}(z') \quad (6)$$

where \bar{w} is the integrated dark matter correlation function (as defined in Eq. 5) and $\bar{b}_u(z)$ and $\bar{b}_r(z)$ are the corresponding integrated biases, defined as the square root of the ratios between the galaxy and dark matter correlation functions. We now consider the case for which $\phi_r(z) \rightarrow \delta_D(z' - z)$. In practice approaching this limit requires selecting a reference population within a redshift slice centered on z with a narrow width δz . If one wishes to detect a non-zero cross-correlation at that redshift, this imposes a lower limit on the density of reference objects in the sky. When focusing on small angular scales, where measurements are limited by shot noise, it is given by

$$\frac{dN_r}{dz} > \frac{1}{\delta z} \frac{\pi \theta^2}{\langle n_u \rangle \bar{w}^2} . \quad (7)$$

As an example, if we consider clustering measurements on scales of 1 degree, this translates into

$$\frac{dN_r}{dz} \gtrsim 10^3 \left(\frac{0.1}{\delta z} \right) \left(\frac{100 \text{ deg}^{-2}}{\langle n_u \rangle} \right) . \quad (8)$$

As we will show in section 2.3, various datasets satisfy this criterion. We note that the detectability condition given in Eq. 7 is not required for redshift inference. The method can be used to generate a large number of noise-dominated cross-correlation estimates which can then be analyzed statistically. In the limit of a narrow reference redshift slice, the redshift distribution of the unknown sample is

$$\phi_u(z) = \bar{w}_{ur}(z) \times \frac{1}{\bar{b}_u(z)} \times \frac{1}{\bar{b}_r(z) \bar{w}(z)} \quad (9)$$

In this equation, the only unknown quantity is $\bar{b}_u(z)$. The integrated dark matter correlation function $\bar{w}(z)$ can be estimated and the redshift-dependent bias of the reference sample can be obtained from a measurement of its auto-correlation function:

$$\bar{w}_{rr}(z) = \bar{b}_r^2(z) \bar{w}(z) . \quad (10)$$

While this relation is valid only on scales where galaxies are linearly biased with respect to the dark matter field, the inclusion of smaller scales provides only a modest departure from it. We demonstrate this point in our companion paper (Schmidt et al. 2013) using numerical simulations. One reason for this is that the scale dependence of the bias is usually a slowly varying quantity. Moreover, our estimate is based on an average over a wide range of scales which weakens the non-linear effects.

It is important to realize that degree of variation of each term in equation 9 is expected to differ. We denote Δz as the range over which $\phi_u(z)$ is greater than zero. If over this range the relative variation of $\phi_u(z)$ dominates over that of $b_u(z)$, or in other words if

$$\frac{d \log \phi_u}{dz} \gg \frac{d \log \bar{b}_u}{dz} \quad (11)$$

then, over the interval Δz , we have

$$\phi_u(z) \propto \bar{w}_{ur}(z) \left(\frac{1}{\bar{b}_r(z) \bar{w}(z)} \right) . \quad (12)$$

The more limited the interval Δz , the larger the variation of $\phi_u(z)$ and the smaller the variation of the other quantities. The proportionality constant depends on the value

of the integrated bias of the unknown sample. However, if we know that all objects of this sample contribute to the spatial cross-correlations with the reference sample then we can simply normalize the redshift distribution using

$$\int dz d\Omega \phi_u(z) = N_u . \quad (13)$$

This relation is typically satisfied if all the objects of the unknown sample are extragalactic *and* if the redshift distribution of the reference population is wide enough to cover the redshift range of all unknown objects. This implies that, in the case of a narrow redshift distribution, it is possible to infer it without the knowledge of the bias of the unknown population. Below, we discuss observational strategies leading to reducing the redshift support Δz .

If the redshift evolution of the bias $b_u(z)$ over Δz is not negligible compared to the variation of $\phi_u(z)$, this method only allows us to estimate the product $\phi_u(z) b_u(z)$. Additional information from the auto-correlation of the unknown sample can be used to attempt breaking the degeneracy between these two quantities (see for example Newman 2008). In the present study we propose to focus on a local sampling of the redshift distribution of an unknown population, as opposed to the methods proposed to infer the global distribution.

In the general case, Eq. 12 provides us with an estimator testing for the absence or existence of sources at a given redshift z , i.e. a *data-driven approach to redshift estimation* which can be applied to any continuous or discrete dataset. When probing sources for which spectral energy distribution templates are not available (for example because the physics of the objects is not understood) or for which no spectroscopic data is available, the cluster-based redshift estimation proposed in this paper provides us with a robust way to infer the presence/absence of sources as a function of redshift, without any assumption.

2.3. Data analysis strategy

The method presented in the previous section is better suited for probing the redshifts of an unknown population for which objects exist only within some limited redshift interval Δz . In practice, sky surveys often provide us with a series of observables for each source (e.g. brightness, colors, size, shape, etc.). In this case, the best approach to the characterization of the redshift distribution of a given dataset is to first select subsamples in the space of all observable parameters. Each subsample will, by construction, live in a redshift interval narrower than that of the entire population. The more parameters are available, the more likely it is to identify regions of that space mapping to narrow redshift intervals. The ability of selecting subsamples in the corresponding multidimensional parameter space is important.

In addition, we note that the higher the redshift sampling given by the reference population, the more likely we are to detect high contrast features in the redshift distribution of the unknown sample. The width δz of the reference redshift slices should be as small as possible. Eq. 7 shows that datasets with more than thousands of objects per unit redshift are required to detect a typical cross-correlation signal. Interestingly we now

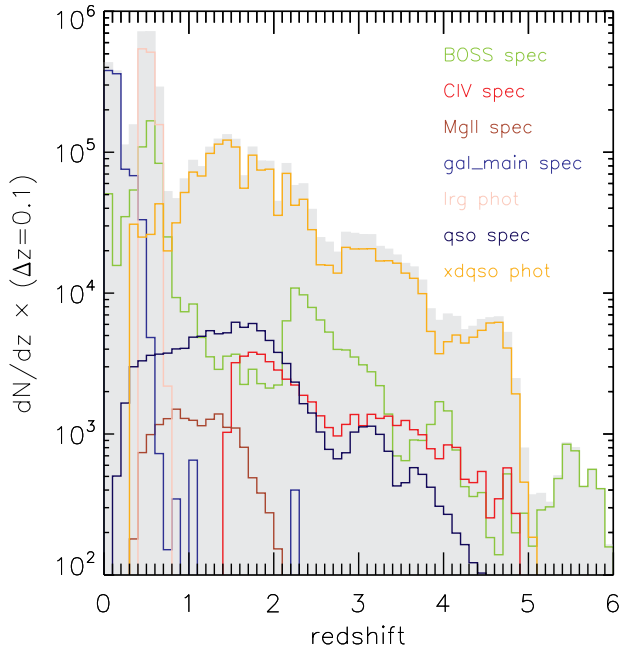


FIG. 1.— Compilation of samples from the SDSS for which we have a robust 3d position, either from spectroscopic or photometric redshifts. In this paper we make use of the spectroscopic samples of quasars and Mg II absorbers as shown with the dark blue and brown curves.

have access to a variety of surveys providing us with 3d positions (based on spectroscopic redshifts or, in some cases, sufficiently accurate photometric redshifts) which are large enough. As an illustration, we show in figure 1 a compilation of samples drawn from the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) for which the redshift distributions are known. The figure includes distributions for galaxies, quasars and absorber systems. As can be seen, the usability criterion given in Eq. 7 is met by numerous samples. This figure also shows that different populations can be used to check the consistency of the inferred redshift distributions.

In the next section we will make use of the spectroscopic quasar and absorber samples as reference populations. Those are shown with the dark blue and brown curves, respectively. While SDSS quasars are found all roughly all redshifts from 0 to 6, Mg II absorbers are only visible in the range $0.4 < z < 2.2$.

2.4. Gravitational lensing effects

The apparent spatial density of sources in the sky is modulated by gravitational magnification effects due to the matter distribution along the line-of-sight (Narayan 1989, e.g.). This induces an apparent correlation between populations of objects lying at different redshifts. The amplitude of this effect, also called cosmic magnification, has been estimated by several authors (see Bartelmann, M. & Schneider, P. 2001) and detected by the large-scale distribution of galaxies by Scranton et al. (2005) and Ménard (2010). For sources at high redshift lensed by typical galaxies at $z \sim 0.5$, the amplitude of the magnification effect is about 1% on a scale of one arcminute. In general, this is negligible compared to the signal induced by physical clustering of overlapping samples. In addition, the redshift dependence of the lensing efficiency

varies slowly with redshift. The absence of such a signature in the redshift distribution inferred by the spatial cross-correlation technique directly indicates that cosmic magnification effects are not playing a significant role.

3. APPLICATION TO DATA

We now apply our method to estimate the redshift distribution of several populations: (i) Luminous Red Galaxies (LRGs) for which accurate photometric redshifts are available for comparison, (ii) Emission Line Galaxies (ELGs) for which photometric redshift estimation is more difficult to estimate due to the presence of strong emission lines, (iii) infrared sources from WISE survey and (iv) radio sources from the FIRST survey, for which photometric redshifts for the single radio flux density are difficult to define. In the first two cases we will use both spectroscopic quasars and Mg II absorbers as reference samples, specifically the SDSS DR7 quasar catalog (Schneider et al. 2010) and the DR7 MgII catalog compiled by Zhu & Ménard (2012). These two samples have different bias evolution profiles, so comparing recoveries on the same unknown sample is a good test that our technique is insensitive to the reference sample’s bias.

We measure spatial cross-correlations between each ‘unknown’ sample and the two spectroscopic populations, integrating over physical scales ranging from zero to 1 Mpc, using a simple weight function $W(r) \propto 1/r$. Our estimator for the redshift distribution $\phi(z)$ is simply normalized according to Eq. 13. Our goal here is not to construct an optimal estimator but to demonstrate that this technique provides us with a new type of information on redshift distributions, independent of what is obtained through photometric redshifts.

When the inferred redshift distribution is broad, we need to take into account the redshift dependence of the bias of the reference population. For these recoveries, we use only our quasar sample, taking our bias evolution from Porciani & Norberg (2006):

$$b_{\text{QSO}}(z) = \frac{1}{\sigma_8} \left[1 + \left(\frac{1+z}{2.5} \right)^\alpha \right] \quad (14)$$

with $\gamma = 4$ to provide a better fit to the high-redshift quasar clustering measurements (Shen et al. 2012).

3.1. Luminous Red Galaxies

We now apply our technique to the MegaZ-LRG sample (Collister et al. 2007). This catalogue contains about one million SDSS Luminous Red Galaxies with robust photometric redshifts. This sample spans the redshift range $0.4 < z < 0.7$ with limiting magnitude $i < 20$. The 2dF-SDSS LRG and Quasar (2SLAQ; Cannon et al. 2006) spectroscopic redshift catalogue of 13 000 intermediate-redshift LRGs provides a photometric redshift training set, indicating that the rms photometric redshift accuracy obtained for an evaluation set selected from the 2SLAQ sample is $\sigma_z = 0.049$ averaged over all galaxies. The distribution of photometric redshifts is shown in Figure 2 with the solid line.

We measure the spatial cross-correlation between LRGs and quasars as a function of redshift, and use it to estimate the LRG redshift distribution. The results are shown with the black data points. They demonstrate that the overall shape of the LRG redshift distribution

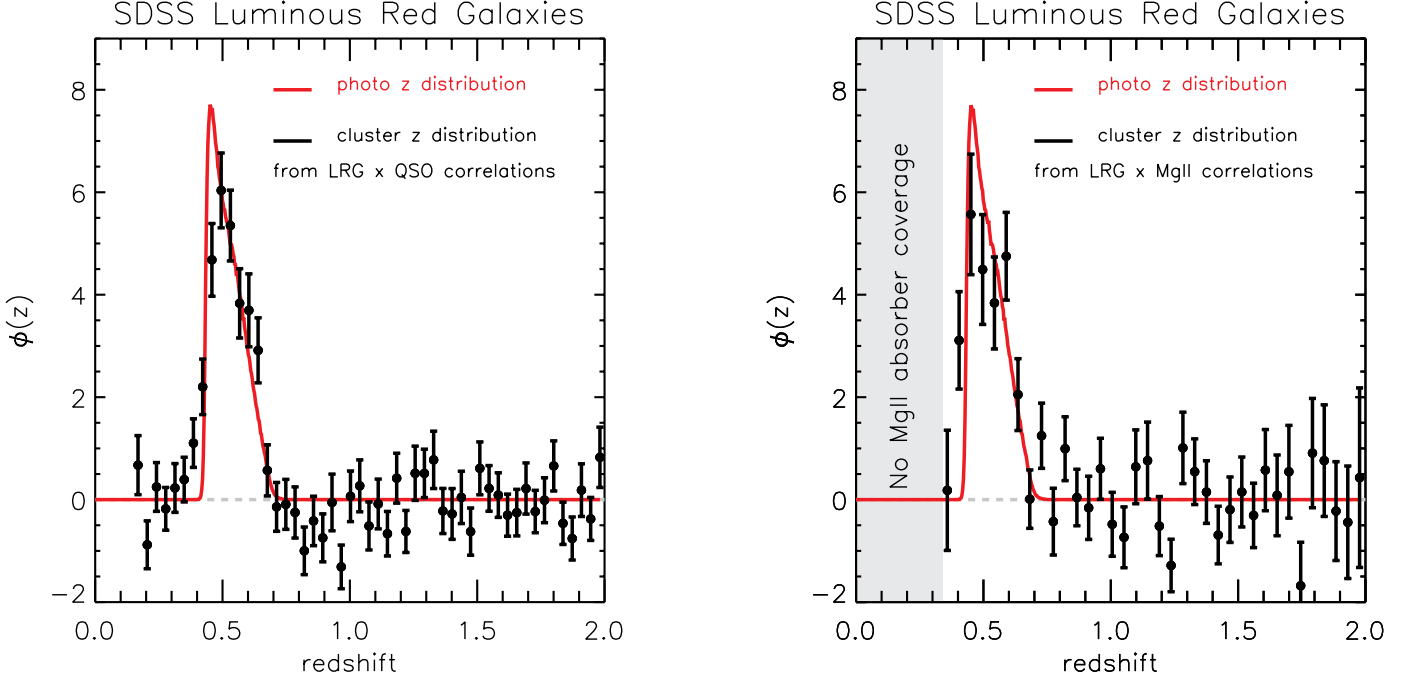


FIG. 2.— Redshift distributions of Luminous Red Galaxies (LRGs). In both panels the solid red line shows the distribution of LRG photometric redshifts. *Left:* cluster- z distribution (black points) obtained by measuring the spatial cross-correlation between LRGs and SDSS quasars. *Right:* cluster- z distribution (black points) obtained by measuring the spatial cross-correlation between LRGs and Mg II absorbers, spanning the range $0.4 < z < 2$.

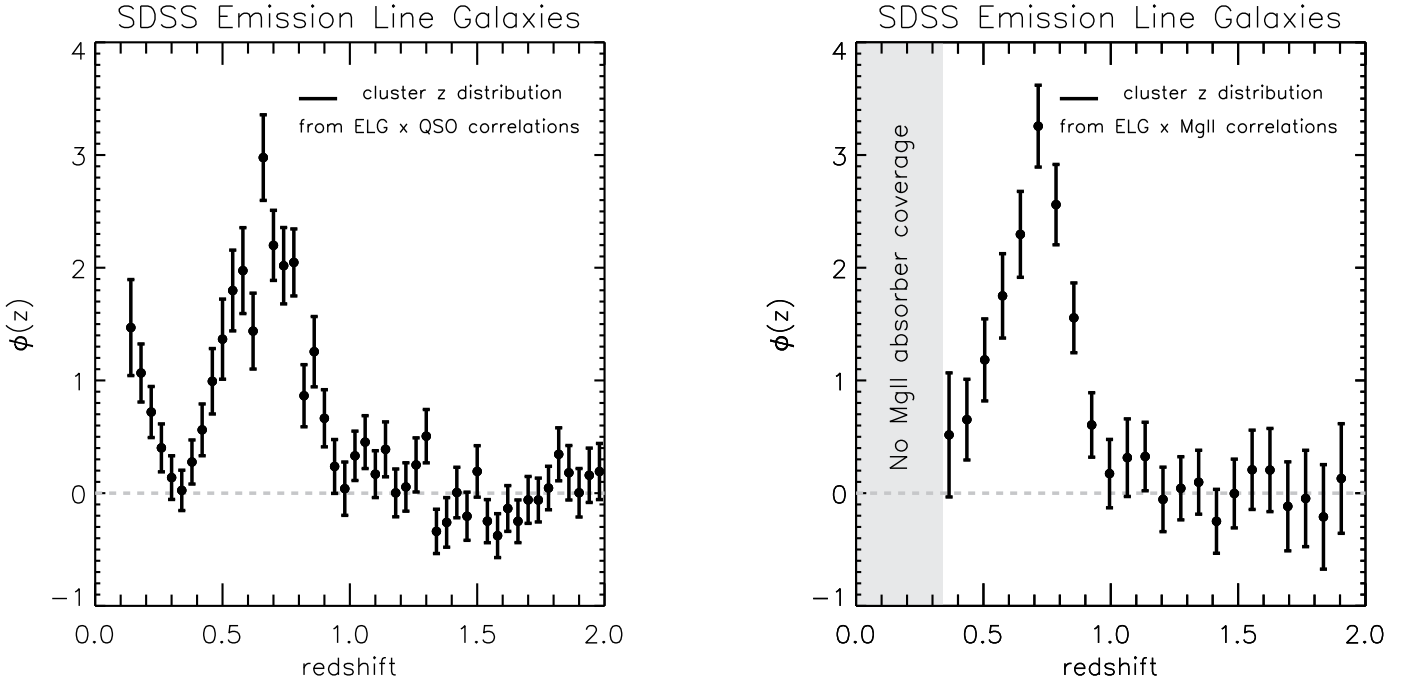


FIG. 3.— Redshift distributions of Emission Line Galaxies (ELGs) from the SDSS. *Left:* cluster- z distribution (black points) obtained by measuring the spatial cross-correlation with SDSS quasars. *Right:* cluster- z distribution (black points) obtained by measuring the spatial cross-correlation with Mg II absorbers, spanning the range $0.4 < z < 2$.

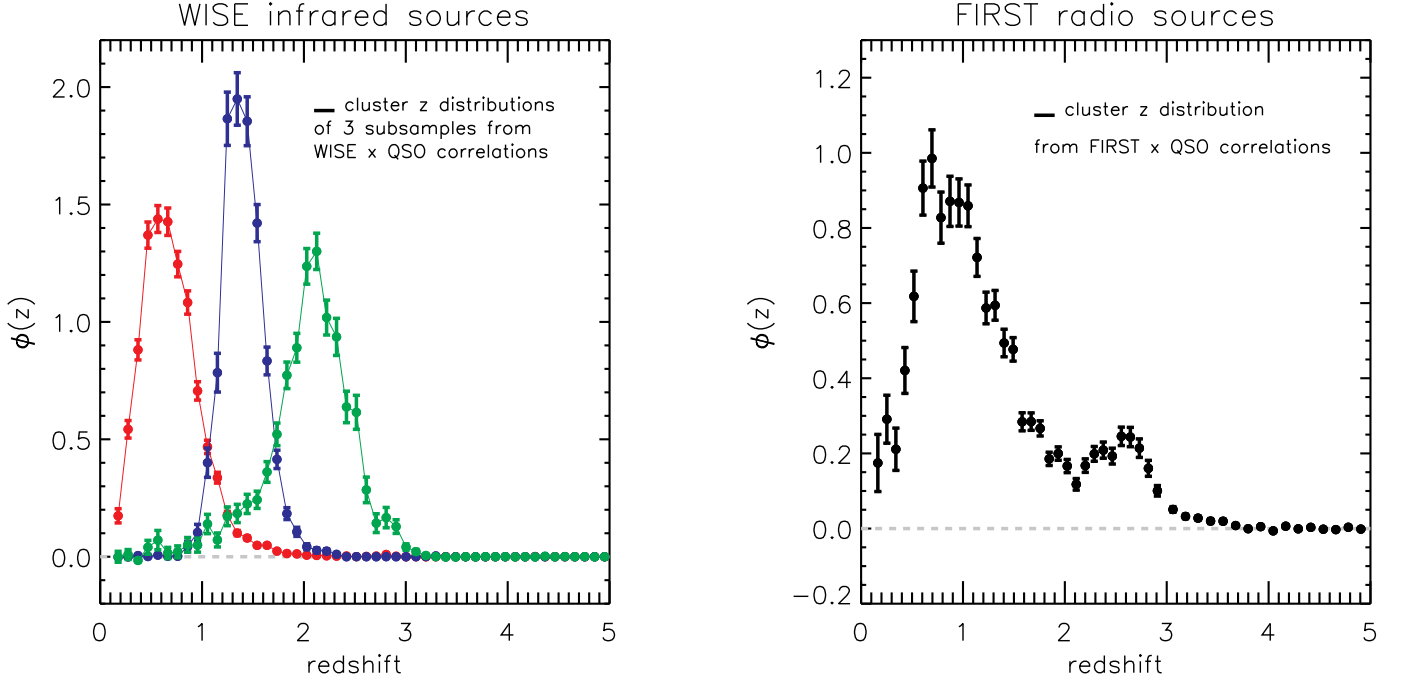


FIG. 4.— *Left*: Redshift distributions of three subsamples of WISE sources obtained by measuring their spatial cross-correlation with SDSS quasars. We show the selection criteria for red (Sample 1), blue (Sample 2) and green (Sample 3) samples in Eq. 16. *Right*: Redshift distribution of FIRST radio sources obtained by measuring their spatial cross-correlation with SDSS quasars. We observe the existence of sources up to $z \sim 3$ as well as a bimodal redshift distribution.

is properly recovered. In addition, the results show that the megaZ-LRG sample is not significantly contaminated by galaxies at other redshifts in the range probed by the quasars.

We then repeat our measurement replacing the quasars with Mg II absorbers. The results, as shown in the right panel of Figure 2, are again in good agreement with the photometric redshift distribution. This provides us with an estimate independent from that obtained with the quasars and shows that different reference samples can be used to obtain consistent results.

3.2. Emission Line Galaxies

We now apply our redshift estimation technique to the so-called Emission Line Galaxies (ELGs) from the SDSS (Comparat et al. 2013). This corresponds to a sample of faint blue galaxies for which the broad band colors are dominated by emission lines. Following these authors, we have selected the galaxies from the SDSS DR7 database with:

$$\begin{aligned} i &< 21.5 \\ g - r &< 1.0 \\ r - i &> -0.917(g - r) + 0.683 \\ r - i &> 0.5(g - r) + 0.4 \end{aligned} \quad (15)$$

Using SDSS DR7, this provides us with a sample of about 2.6 million galaxies. We measure the spatial cross-correlation between these ELGs and quasars as a function of redshift and use it to estimate the redshift distribution of the population. The results are shown in Figure 3 with the black data points. They indicate that the ELG redshift distribution is bi-modal, with a main population located at $z \sim 0.6$ and a second group located

at lower redshift.

We also measure the spatial cross-correlation between ELGs and Mg II absorbers as a function of redshift. Again, the recovered redshift distribution is in good agreement with that obtained from the spectroscopic quasars. In this case, the overall normalization given by Eq. 13 does not properly apply as the spectroscopic redshift coverage is not wide enough to probe the redshifts of all unknown sources. As a result, the amplitude of $\phi_u(z \sim 0.7)$ obtained with the Mg II absorber systems is higher than that the more correct one obtained with quasars as the reference population.

Because the redshift distribution is not simple and we are most likely observing two distinction populations of galaxies with different biases, we cannot make any strong claims about the relative numbers of the low and high redshift populations. However, with the redshift recovery technique at our disposal, we are not limited to accepting these results as final. By iterating between selection cuts and recoveries, we could, for instance, tune the selection criteria in Equation 15 to exclude the low redshift population.

3.3. The WISE infrared survey

The Wide-Field Infrared Survey Explorer (WISE; Wright et al. 2010) is a mid-infrared survey satellite which provides us with all-sky observations in four bands, centered at 3.4, 4.6, 12, and 22 μm (W1 to W4, hereafter). In order to maintain homogeneous WISE sample, we first select WISE sources with magnitude cut $[W_1] < 16.5$.

For illustration purposes we work with three subsamples selected in color-color space, $([W_2] - [W_3]) - ([W_1] -$

$[W_2])$:

$$\begin{aligned}
 \text{Sample 1 : } & 2 < [W_{2-3}] < 2.5 \\
 & 0.9 < [W_{1-2}] < 1.2 \\
 \text{Sample 2 : } & 2.5 < [W_{2-3}] < 3 \\
 & 1.5 < [W_{1-2}] < 1.8 \\
 \text{Sample 3 : } & 3.5 < [W_{2-3}] < 4 \\
 & 1.2 < [W_{1-2}] < 1.5
 \end{aligned} \tag{16}$$

where $[W_{i-j}] = [W_i] - [W_j]$. We then cross-correlate these subsamples against the SDSS QSOs. As a result, we find the clear trend with redshift. In Figure 4, we present the redshift distributions of these three subsamples obtained by cross-correlations with QSOs as our reference sample. In Figure 4, the three samples are shown with different colors: Sample 1 (red), Sample 2 (blue) and Sample 3 (green). While these samples represent only a small fraction of the WISE data, they show that even simple color cuts may be sufficient for selecting non-overlapping samples for cosmological tests (e.g. lensing). A future paper will explore the redshift distribution of the WISE data in more detail.

3.4. The FIRST radio survey

The Faint Images of the Radio Sky at Twenty cm survey (FIRST; Becker et al. 1995) uses the Very Large Array (VLA) to produce a map of the 20 cm (1.4 GHz) sky with a beam size of $5.4''$ and an rms sensitivity of about 0.15 mJy/beam. The survey covers an area of about $10,000 \text{ deg}^2$ in the north Galactic cap and a smaller area along the celestial equator, both of which roughly coincide with the regions observed by SDSS. With a source surface density of $\sim 90 \text{ deg}^{-2}$, the final catalog includes about one million objects.

Using our spectroscopic quasar catalog and correcting for bias evolution as given in Equation 14, we recovered the redshift distribution shown in Figure 4. As mentioned in §2.2, this is a broad redshift distribution where we are violating our conditions from Equation 11. Hence, we do not expect that our recovery is independent from evolving bias in the FIRST sample. However our results allow us to say with some confidence that the source redshift distribution extends to $z \sim 3$ and that there exists two distinct populations of sources, one centered around $z \sim 1$ and a higher redshift cohort around $z \sim 2.5$. Selecting these two populations independently is difficult from radio data only, given the lack of additional parameters available in FIRST, but can be done via cross-matching FIRST sources with external datasets (Schmidt et al., in preparation).

4. CONCLUSIONS

We have presented a method to infer the redshift distribution of arbitrary datasets, based on spatial cross-correlations with a reference population and we have applied it to various datasets across the electromagnetic spectrum. Previous works exploring the same avenue

(e.g. Newman 2008; Matthews & Newman 2010; Schulz 2010; Matthews & Newman 2012; McQuinn & White 2013) have focused on large scales where the halo bias is linear. Here we advocate the use of clustering measurements on all available scales and discuss the benefits of using small-scale correlations which tend to be less affected by systematics with real data. In a companion paper (Schmidt et al. 2013) we have used numerical simulations to show the robustness and limitations of this approach. We have also used this technique to search for contamination of high redshift Lyman-break galaxies by low redshift interlopers (Morrison et al. 2012).

Here, we have applied our method to estimate the redshift distributions of SDSS luminous red galaxies, emission line galaxies, sources from the WISE infrared survey and the FIRST radio survey. For the first two samples, located at low redshift, we have estimated their redshift distributions using both quasars and absorber systems as the reference population and obtained consistent results. The simple, narrow redshift distributions recovered for the LRGs and WISE sub-samples should be reliable high S/N estimates of the underlying redshift distributions for these samples. For the broader, multi-peaked distributions recovered for the ELG and FIRST samples, we do not expect our recoveries to be unbiased estimates of the distributions, but we are able to make reliable claims about the redshifts of the sub-populations contained in these samples. Additionally, an iterative approach combining sample selection and redshift recovery has the potential to greatly aid in increasing the purity of these samples, separating low and high redshift populations.

This technique promises to be widely applicable to existing and upcoming sky surveys. It provides us with the invaluable ability to deproject the inherently 2d observations of the extragalactic sky.

This work is supported by a NASA grant and the Alfred P. Sloan foundation. RS, SJS and CBM acknowledge the support of NSF Grant AST-1009514. DJ acknowledges the support of DoE SC-0008108 and NASA NNX12AE86G.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

REFERENCES

- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
- Bartelmann, M., Schneider, P. 2001, Phys. Rep., 340, 291
- Becker, R. H. and White, R. L. and Helfand, D. J. 1995, ApJ, 450, 559
- Benjamin J., van Waerbeke L., Ménard B., Kilbinger M., 2010, MNRAS, 408, 1168

- Cannon, R., Drinkwater, M., Edge, A., et al. 2006, MNRAS, 372, 425
- Collister, A., Lahav, O., Blake, C., et al. 2007, MNRAS, 375, 68
- Comparat, J., Kneib, J.-P., Escoffier, S., et al. 2013, MNRAS, 428, 1498
- Matthews D. J., Newman J. A., 2010, ApJ, 721, 456
- Matthews D. J., Newman J. A., 2012, ApJ, 745, 180
- McQuinn, M., & White, M 2013, arXiv:1302.0857
- Ménard, B., Scranton, R., Fukugita, M., Richards,, G. 2010, MNRAS, 405, 1025
- Morrison, C. B., Scranton, R., Ménard, B., et al. 2012, MNRAS, 426, 2489
- Narayan, R. 1989, ApJ, 339, L53
- Newman J. A., 2008, ApJ, 684, 88
- Peebles, P. J. E. 1993, Principles of Physical Cosmology by P.J.E. Peebles. Princeton University Press, 1993. ISBN: 978-0-691-01933-8,
- Porciani, C., & Norberg, P. 2006, MNRAS, 371, 1824
- Schmidt, S., Ménard, B., Scranton, R., Morrison, C., & McBride, C. 2013, arXiv:1303.0292
- Schulz A. E., 2010, ApJ, 724, 1305
- Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360
- Scranton, R., Ménard, B., Richards, G. T., et al. 2005, ApJ, 633, 589
- Shen, Y., McBride, C. K., White, M., et al. 2012, arXiv:1212.4526
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
- Zhu, G., Ménard, B. 2012, arXiv:1211.6215